

Comparing two thesaurus representations for Russian

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2018 Global WordNet Association. All Rights Reserved. In the paper we presented a new Russian wordnet, RuWordNet, which was semi-automatically obtained by transformation of the existing Russian thesaurus RuThes. At the first step, the basic structure of wordnets was reproduced: synsets' hierarchy for each part of speech and the basic set of relations between synsets (hyponym-hypernym, part-whole, antonyms). At the second stage, we added causation, entailment and domain relations between synsets. Also derivation relations were established for single words and the component structure for phrases included in RuWordNet. The described procedure of transformation highlights the specific features of each type of thesaurus representations.

References

- [1] Saima Aman and Stan Szpakowicz. 2008. Using Roget's Thesaurus for Fine-grained Emotion Recognition, Proceedings of IJCNL-2008: 312-318.
- [2] Irina Azarova. 2008. RussNet as a Computer Lexicon for Russian, Proceedings of the Intelligent Information systems IIS-2008: 341-350.
- [3] Irina Azarova, Pavel Braslavski, Viktor Zakharov, Yuri Kiselev, Dmitrii Ustalov and Maria Khohlo-va. 2016. Integration of thesauri RussNet и YARN. Proceedings of "Internet and Modern Society" conference IMS-2016 (in Russian).
- [4] Valentina Balkova, Andrey Suhonogov, and Sergey Yablonsky. 2008. Some Issues in the Construction of a Russian WordNet Grid. In Proceedings of the Forth International WordNet Conference, Szeged, Hungary:44-55.
- [5] Luisa Bentivogli, Pamela Forner, Bernardo Magnini, B., and Emanuele Pianta. 2004. Revising the wordnet domains hierarchy: semantics, coverage and balancing. In Proceedings of the Workshop on Multilingual Linguistic Resources: 101-108.
- [6] Francis Bond, and Paik Kyonghee. 2012. A survey of wordnets and their licenses. Proceedings of Global WordNet Conference GWC-2012: 64-71.
- [7] Lars Borin and Markus Forsberg. 2009. All in the family: A comparison of SALDO and WordNet. Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources, Odense.
- [8] Lars Borin, Markus Forsberg and Lennart Lonngrén. 2013. SALDO: a touch of yin to WordNet's yang. Language Resources and Evaluation, 47(4):1191-1211.
- [9] Lars Borin, Jens Allwood, and Gerard de Melo. 2014 .Bring vs. MTRoget: Evaluating automatic thesaurus translation. Proceedings of LREC-2014, Reykjavik, ELRA: 2115-2121.
- [10] Pavel Braslavski, Dmitrii Ustalov and Mikhail Mukhin. 2014, A Spinning Wheel for Yarn: User Interface for a Crowdsourced Thesaurus, Proceedings of EACL-2014, Gothenberg, Sweden: 101-104.
- [11] Boris Dobrov and Natalia Loukachevitch. 2006. In Development of Linguistic Ontology on Natural Sciences and Technology. In Proceedings of LREC-2006.
- [12] Boris Dobrov and Andrey Pavlov. 2010. A Basic line for news clusterization methods evaluation. Proceedings of RCDL-2010: 287-295.
- [13] Christiane Fellbaum. 1998. WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

- [14] Alfiya Galieva, Olga Nevzorova and Dilyara Yaku-bova . .2017. Russian-Tatar Socio-Political Thesaurus: Methodology, Challenges, the Status of the Project. In Proceedings of Recent Advances in Natural Language Processing Conference (RANLP-2017): 245-252
- [15] Ilia Gelfenbeyn, Artem Goncharuk, Vlad Lehelt, Anton Lipatov, and Viktor Shilo. 2003. Automatic translation of WordNet semantic network to Russian language. In Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2003.
- [16] Derek Gross and Katherine Miller. 1990. Adjectives in WordNet, *International Journal of Lexicography*, 3(4):.265-277.
- [17] Nicola Guarino. 2009. The Ontological Level: Revisiting 30 Years of Knowledge Representation. In *Conceptual Modeling: Foundations and Applications: Essays in Honor of John Mylopoulos*. Lecture Notes in Computer Science, 5600, Berlin and Heidelberg, Germany: Springer-Verlag: 52-67.
- [18] Mario Jarmasz and Stan Szpakowicz. 2004. Roget's thesaurus and semantic similarity, *Recent Advances in Natural Language Processing III: Selected Papers from RANLP, 2004*: 111-120.
- [19] Alistair Kennedy, and Stan Szpakowicz. 2008. Evaluating Roget's Thesauri, *Proceedings of ACL-2008*: 416-424.
- [20] Mikhail Khodak, Andrej Risteski, Christiane Fellbaum, and Sanjeev Arora. 2017. Automated WordNet Construction Using Word Embeddings. *Proceedings of SENSE 2017*: 12-23.
- [21] Claudia Kunze and Lothar Lemnitzer. 2010. Lexical-Semantic and Conceptual relations in GermaNet. In Storjohann P (ed) *Lexical-semantic relations: Theoretical and practical perspectives*, 28:163-183.
- [22] Svetlozara Leseva, Maria Todorova, Tsvetlana Dimi-trova, Borislav Rizov, Ivelina Stoyanova, Svetla Koeva. 2015. Automatic classification of wordnet morphosemantic relations. In *The 5th Workshop on Balto-Slavic Natural Language Processing*: 59-64.
- [23] Krister Lindén and Jyrki Niemi. 2014. Is it possible to create a very large wordnet in 100 days? An evaluation, *Language resources and evaluation*, 48.2:191-201.
- [24] Natalia Loukachevitch and Boris Dobrov. 2002. Development and Use of Thesaurus of Russian Language RuThes. *Proceedings of workshop on WordNet Structures and Standartisation, and How These Affect WordNet Applications and Evaluation*. LREC-2002: 65-70.
- [25] Natalia Loukachevitch and Boris Dobrov. 2014. RuThes Linguistic Ontology vs. Russian Wordnets. *Proceedings of Global WordNet Conference GWC-2014*, Tartu: 154-162.
- [26] Natalia Loukachevitch, Boris Dobrov and Ilia Chetvi-orkin. 2014. RuThes-Lite, a publicly available version of thesaurus of Russian language RuThes. In *Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2014*, 340-350.
- [27] Natalia Loukachevitch and Aleksey Alekseev. 2014. Summarizing News Clusters on the Basis of Thematic Chains. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC-2014)*.
- [28] Natalia Loukachevitch and Boris Dobrov. 2015. The Sociopolitical Thesaurus as a resource for automatic document processing in Russian. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication* 21.2: 237-262.
- [29] Natalia Loukachevitch and German Lashevich. 2016. Multiword expressions in Russian thesauri RuThes and RuWordNet. In *Artificial Intelligence and Natural Language Conference (AINL-2016)*, IEEE: 1-6.
- [30] Natalia Loukachevitch, Alexander Shevelev and Valeria Mozharova. 2017. Testing Features and Methods in Russian Paraphrasing Task. In *Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2017*, V.1.: 135-146.
- [31] Natalia Loukachevitch and Anastasia Gerasimova. 2017. Human Associations Help to Detect Conventionalized Multiword Expressions. *arXiv preprint arXiv:1709.03925*.
- [32] Bernardo Magnini and Gabriela Cavaglia. 2000. Integrating Subject Field Codes into WordNet. In *Proceedings of Language Resources and Evaluation Conference LREC-2000*: 1413-1418.
- [33] Inderjeet Mani, I., Klein, G., House, D., Hirschman, L., Firmin, T., and Sundheim, B. (2002). SUM-MAC: a text summarization evaluation. *Natural Language Engineering*, 8 (1), 43-68.
- [34] Marek Maziarz, Stanoslaw Szpakowicz and Maciej Piasecki. 2012. Semantic relations among adjectives in Polish WordNet 2.0: a new relation set, discussion and evaluation. *Cognitive Studies*, 12:,149-179.
- [35] George Miller. 1998. Nouns in WordNet. In *WordNet - An Electronic Lexical Database*, edited by Christiane Fellbaum, Cambridge, MA: The MIT Press: 23-47.
- [36] George Miller and Florentina Hristea. 2006. WordNet Nouns: Classes and Instances. *Journal of Computational linguistics*, 32(1):1-3.
- [37] Michael Nokel and Natalia Loukachevitch. 2016. Accounting ngrams and multi-word terms can improve topic models. In *Proceedings of 12th Workshop on Multiword Expressions, ACL 2016*: 44-49.
- [38] Karel Pala and Dana Hlaváčková. 2007. Derivational relations in czech wordnet. *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing: Information Extraction and Enabling Technologies*. Association for Computational Linguistics.

- [39] Bolette Pedersen, Sanni Nimb, Jorg Asmussen, Nico-lai Sørensen, Lars Trap-Jensen L and Henrik Lo-rentzen. 2009. DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary, *Language resources and evaluation*, 43(3): 269-299.
- [40] Maciej Piasecki, Radoslaw Ramocki, and Marek Maziarz. 2012. Automated generation of derivative relations in the Wordnet expansion perspective. In *Proceedings of the 6th Global Wordnet Conference GWC 2012*: 273-280.
- [41] Piek Vossen. 1998. Introduction to EuroWordNet. In *EuroWordNet: A multilingual database with lexical semantic networks*, Springer Netherlands: 1-17.
- [42] Z39.19. 2005. Guidelines for the Construction, Format and Management of Monolingual Thesauri. NISO.